





ORIGINAL ARTICLE

Perilesional sun damage as a diagnostic clue for pigmented actinic keratosis and Bowen's disease

P. Weber,^{1,†} C. Sinz,^{1,†}  C. Rinner,²  H. Kittler,¹  P. Tschandl^{1,*} 

¹Department of Dermatology, Medical University of Vienna, Vienna, Austria

²Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

*Correspondence: P. Tschandl. E-mail: philipp.tschandl@meduniwien.ac.at

Abstract

Background Chronic sun damage in the background is common in pigmented actinic keratoses and Bowen's disease (pAK/BD). While explainable artificial intelligence (AI) demonstrated increased background attention for pAK/BD, humans frequently miss this clue in dermatoscopic images because they tend to focus on the lesion.

Aim To analyse whether perilesional sun damage is a robust diagnostic clue for pAK/BD and if teaching this clue to dermatoscopy users improves their diagnostic accuracy.

Methods We assessed the interrater agreement and the frequency of perilesional sun damage in 220 dermatoscopic images and conducted a reader study with 124 dermatoscopy users. The readers were randomly assigned to one of two online tutorials; one tutorial pointed to perilesional sun damage as a clue to pAK/BD (group A) the other did not (group B). In both groups, we compared the frequencies of correct diagnoses before and after receiving the tutorial.

Results The frequency of perilesional sun damage was higher in pAK/BD than in other types of pigmented skin lesions and interrater agreement was good ($\kappa = 0.675$). The diagnostic accuracy for pAK/BD improved in both groups of readers (group A: +16.1%, 95%-CI: 9.5–22.7; group B: +13.1%; 95%-CI: 7.1–19.0; P for both <0.001), but the overall accuracy improved only in group A from (59.1% (95%-CI: 55.0–63.1) to 63.5% (95%-CI: 59.5–67.6); $P = 0.002$).

Conclusion Perilesional sun damage is a good clue to differentiate pAK/BD from other pigmented skin lesions in dermatoscopic images, which could be useful for teledermatology. Knowledge of this clue improves the accuracy of dermatoscopy users, which demonstrates that insights from explainable AI can be used to train humans.

Received: 15 March 2021; Accepted: 10 June 2021

Conflicts of interest

PW reports funding from the International Dermoscopy Society (2019 IDS Research Grant) supporting the present manuscript. CS reports no conflict of interest. CR reports grants from Lilly, outside the submitted work. HK reports grants from Lilly, honoraria from Lilly, Novartis, FotoFinder and Pelpharma and received equipment from Heine, FotoFinder, DermaMedical, outside the submitted work. PT reports grants from MetaOptima and Lilly, consulting fees from Silver-chair and honoraria from Lilly, outside the submitted work.

Funding sources

International Dermoscopy Society (2019 IDS Research Grant).

Introduction

Actinic keratosis (AK) and Bowen's disease (BD), which is also referred to as intraepithelial carcinoma (IEC), are regarded as non-invasive variants of squamous cell carcinoma of the skin (SCC).¹ Histologically, BD is characterized by dysplasia of the entire epidermis, whereas AK is typified by dysplasia of the basal layers.² Pigmented variants of AK and BD (pAK/BD) are difficult to differentiate from other pigmented skin lesions, notably from lentigo maligna and lichen planus-like keratosis (LPLK), even by

dermatoscopy. A handful of dermatoscopic criteria have been described for pAK/BD but it is unknown if knowledge of these criteria improves the diagnostic accuracy of clinicians.^{3,4}

A case series published in 2010 investigated 52 lesions of pBD and established dermatoscopic criteria based on pattern analysis⁵ such as structureless brown areas, surface scales and pigmented dots or coiled vessels arranged in lines. Another study performed by Yi Yang et al in 2017 described scales, yellow crusts and coiled ('glomerular') vessels as typical characteristics of pBD in 146 individuals of Asian descent.⁶ In comparison to BD, pAK occurs more frequently on the face and is typified by scale, prominent

[†]Contributed equally.

follicular openings, erythema and white circles, which appear as four white dots (rosettes), under polarized dermatoscopy.

Despite various further efforts by different groups to improve the diagnostic accuracy for pAK/BD,^{1,7,8} a recent study demonstrated that human readers are inferior to computer algorithms in a simulated telemedical setting.⁹ While the best computer algorithms diagnosed 90% of pAK/BD correctly, human readers reached the correct diagnosis in only 50%. One reason for this striking human underperformance may be that humans focus on the lesion and not on the background, which in the case of pAK/BD may contain diagnostic information. Except in the rare cases of BD induced by human papillomavirus, perilesional sun damage is always present in pAK/BD and may serve as an additional diagnostic clue. It seems that, in contrast to humans, state-of-the-art computer algorithms harness this bit of information. Bissoto et al., for example, showed that automated image analysis retains high accuracy even when the lesion area is covered,¹⁰ and in a recent study, we demonstrated that background attention of a state-of-the-art computer algorithm is higher in pAK/BD than in any other category of pigmented skin lesions.¹¹

While chronic sun damage is readily assessed in face-to-face consultations, it may be difficult for humans to extract this information from close-up or dermatoscopic images in telemedical consultations or online reader studies. The aims of this study were to estimate the frequency of perilesional sun damage in dermatoscopic images of pAK/BD, to measure the interrater agreement in assessing perilesional sun damage, to assess the efficacy of a short online tutorial on pAK/BD, and to calculate the additional benefit of pointing out the clue of perilesional sun damage.

Methods

To verify if perilesional sun damage is a robust and relevant visual feature, we randomly selected dermatoscopic images of 120 pAK/BD, 20 basal cell carcinomas (BCC), 20 BKL (Benign keratotic lesions: Seborrheic keratosis, solar lentigo and lichen

planus-like keratosis), 20 dermatofibromas, 20 melanomas and 20 melanocytic nevi from the HAM10000¹² dataset. To certify that the raters are not affected by their diagnosis, the centre area (300 × 300 pixels) was filled with black pixels to cover the lesion of interest (Fig. 1). Two raters (H.K., P.T.) separately rated every image for the presence or absence of perilesional sun damage in a single session. Interrater agreement was measured via unweighted Cohen's Kappa. In subsequent analyses, perilesional sun damage was deemed to be present only if both raters agreed. We used a multivariable logistic regression including sex, age and anatomic site as independent variables to estimate the predictive value of perilesional sun damage for pAK/BD (Table 1).

To assess the impact of teaching the clue of perilesional sun damage to dermatoscopists, we designed a reader study and hosted it on an online teaching platform (DermaChallenge, <https://dermonaut.meduniwien.ac.at/dermachallenge>). The readers were recruited via social media channels of the International Dermoscopy Society (IDS). We mainly used the IDS Facebook site for specific calls for this study. While the IDS Facebook site officially has more than 25 000 members, single posts typically reach no more than 5000 IDS members. The exact number of members who received the invitation is unknown. The task of the readers was to select the correct answer from 7 choices (pAK/BD, BCC, BKL, DF, MEL, NV, VASC) for 10 images. The images were randomly selected from all images that were previously rated by the same readers in an otherwise identical quiz but before they received a tutorial. Then the readers were randomly assigned to one of two online tutorials. Both tutorials explained clues for pAK/BD such as angulated lines, prominent follicular openings, white circles, rosettes, erythema, scaling and pigmented dots and coiled vessels in a linear arrangement. The tutorials were identical, except that one tutorial pointed out the additional clue of perilesional sun damage (group A) and the other did not (group B). Users were sequentially assigned to group A and group B in an alternate fashion based on the time

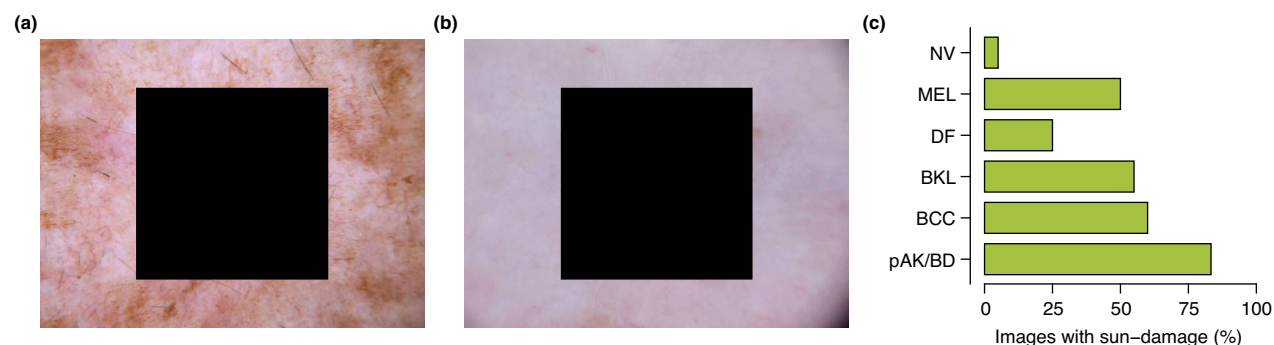


Figure 1 Example images with (a; pAK/BD; ISIC_0024575) and without perilesional sun damage (b; Melanoma; ISIC_0025835). (c) While perilesional sun damage was present in many diagnostic groups, it was the most common in the pAK/BD class.

Table 1 Adjusted odds ratios for predicting pAK/BD from meta-data and sun damage as the only visual feature from a random subset of lesions ($n = 220$)

	OR (95%-CI)	P-value
Sun damage (Consensus)	5.56 (2.82–11.34)	<0.001
Age	1.02 (1–1.05)	0.110
Anatomic site		
Head & neck	5.86 (1.7–22.73)	0.007
Lower extremity	2.12 (0.61–8.2)	0.253
Torso	2.63 (0.62–12.21)	0.199
Upper extremity	3.87 (1.12–14.99)	0.038
Sex Male	1.03 (0.52–2.02)	0.922

Bold values denote those with a p-value <0.05

of enrolment, resulting in two equally sized groups. Each reader was allowed to play a maximum of three iterations of the quiz, each iteration composed of a new set of images. In both groups, we measured the proportion of correct diagnoses before and after watching the tutorial.⁹

The study had a power of >0.9 to detect a 5% increase in the frequency of correct answers between tests taken before and after the intervention within a group. Given the observed effect size (Cohen's d 0.392), the study has a power of 0.58 at a significance level of 0.05 to detect differences between the two groups.

To measure differences between groups, we compared continuous measures with paired or unpaired t -tests or the Mann–Whitney U -Test, as appropriate. To adjust for potential confounders, we applied a multivariate regression model without stepwise elimination, including experience, profession, gender, age and intervention group as variables. We used R-statistics v4.0.3¹³ for all analyses. All P -values are two-tailed and a P -value of <0.05 was regarded as statistically significant. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee of the Medical University of Vienna (EK Nr. 1503/2018).

Results

Frequency and interrater agreement of perilesional sun damage

There was substantial agreement ($\kappa = 0.675$) between two raters who assessed perilesional sun damage independently. Taking the consensus of the two raters as ground truth, perilesional sun damage was least common in nevi (5%, $n = 1$ of 20) and dermatofibromas (25%, $n = 5$ of 20). It was present in 50% of melanomas ($n = 10$ of 20), 55% of benign keratosis lesions ($n = 11$ of 20) and 60% of basal cell carcinomas ($n = 12$ of 20), and was most common in pAK/BD (83.3%, $n = 100$ of 120; Fig. 1).

Reader study

Between January 15th and July 16th 2020, 124 readers (47.6% female) participated in the online study. The readers played 2

rounds on average resulting in 278 reading sets. The mean time span between the first (before the tutorial) and the second reading set (after the tutorial) was 188.6 days (95% CI: 159.3–217.8). We found no relevant differences with regard to the distributions of reader characteristics between group A, which was assigned to the tutorial that pointed out the clue of perilesional sun damage and group B, which was assigned to the other tutorial (Table 2). Both groups spent a similar amount of time on the online tutorial (group A: median 49.19 s (IQR: 33.49) vs. group B: 56.48 (IQR: 37.49), $P = 0.335$). In group A, the mean number of correct answers improved from 59.1% (95% CI: 55.0–63.1) before the tutorial to 63.5% (95% CI: 59.5–67.6; $P = 0.002$) after the tutorial. The number of correct answers did not improve significantly in group B (59.5% (95% CI: 55.9–63.1) vs. 60.3% (95% CI: 56.7–63.8); $P = 0.44$). In a direct comparison, improvement of overall correct answers per user was higher in group A than group B ($P = 0.031$), and in a multivariate model assignment to group A remained an independent predictor of improving previously incorrect answers (OR 1.23; 95%-CI 1.06–1.42; $P = 0.005$). Analysing the results by diagnostic category revealed that both groups improved significantly for pAK/BD (group A: +16.1% (95% CI: 9.5–22.7), $P < 0.001$;

Table 2 Characteristics of readers of group A (assigned to the tutorial pointing to perilesional sun damage) and group B

	Group A	Group B
<i>n</i>	62	62
Age group		
20–30 years	24.2% ($n = 15$)	43.5% ($n = 27$)
31–41 years	41.9% ($n = 26$)	17.8% ($n = 11$)
41–50 years	19.4% ($n = 12$)	16.1% ($n = 10$)
51–60 years	11.3% ($n = 7$)	16.1% ($n = 10$)
61+ years	3.2% ($n = 2$)	6.5% ($n = 4$)
Experience		
<1 year	25.8% ($n = 16$)	38.7% ($n = 24$)
>1 year	19.4% ($n = 12$)	11.3% ($n = 7$)
>3 years	22.6% ($n = 14$)	17.8% ($n = 11$)
>5 years	17.8% ($n = 11$)	19.4% ($n = 12$)
>10 years	14.5% ($n = 9$)	12.9% ($n = 8$)
Gender		
Female	51.6% ($n = 32$)	43.5% ($n = 27$)
Male	48.4% ($n = 30$)	54.8% ($n = 34$)
Other	–	1.6% ($n = 1$)
Profession		
Board certified dermatologist	50.0% ($n = 31$)	35.5% ($n = 22$)
Dermatology resident	21.0% ($n = 13$)	21.0% ($n = 13$)
General practitioner	8.1% ($n = 5$)	17.7% ($n = 11$)
Medical specialist	1.6% ($n = 1$)	–
Medical student	12.9% ($n = 8$)	24.2% ($n = 15$)
Non medical	1.6% ($n = 1$)	–
Nurse practitioner	3.2% ($n = 2$)	1.6% ($n = 1$)
Resident	1.6% ($n = 1$)	–

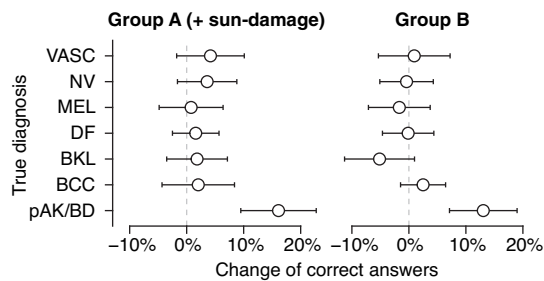


Figure 2 Mean change of correct answers of participants after seeing the short tutorial on pAK/BD clues. Group A was assigned to the tutorial that pointed to perilesional sun damage as an additional clue. Circles denote means, and error bars 95% confidence intervals.

group B: +13.1% (95% CI: 7.1–19.0), $P < 0.001$), but in group B the improvement was outweighed by a decreased diagnostic accuracy for BKL (Figure 2).

Discussion

In this study, we demonstrate that perilesional sun damage is a robust and relevant diagnostic clue for pAK/BD. This clue was detected by the methods of explainable artificial intelligence (AI) in the context of a human–computer collaboration study.¹¹ AI-based decision support may augment human expertise in image-based diagnostic medicine.^{14,15} There are, however, concerns that AI may replace physicians instead of supporting them. Instead of playing off human intelligence against artificial intelligence, it might be better to take the best of both worlds. In a previous study, we showed that perilesional sun damage helps medical students, who had no experience in dermatoscopy, to detect pAK/BD with higher accuracy.¹¹ Here, we demonstrate that teaching this clue improves the accuracy of human dermatoscopy users with varying expertise.

Human–computer collaboration has finally come full-circle: Humans collected and provided training and test images,¹² the accuracy of AI was superior to humans in a subset of cases,⁹ insights of explainable AI were translated into an understandable concept,¹¹ and finally, the concept was used to train humans and improve their diagnostic accuracy.

Furthermore, we demonstrated that minimalistic online tutorials are sufficient to train dermatoscopy users. Both groups improved significantly with regard to the diagnostic accuracy of pAK/BD after seeing a short online tutorial on the clues of pAK/BD. There was no comparable improvement for other diagnoses, which points to the effect of the specific content of the tutorial. However, the overall diagnostic accuracy improved only in the group assigned to the tutorial that pointed out the clue of perilesional sun damage (group A), but not in the other group (group B), which received an otherwise identical tutorial. The reason for this was most likely that in group B the combined decrease of

the accuracy for all other diagnostic categories outweighed the gain in the accuracy for pAK/BD. This effect was most pronounced in the category of benign keratinocytic lesions. Lichen planus-like keratosis (LPLK) in particular is difficult to differentiate from pAK/BD, but our data indicate that this may be easier if perilesional sun damage is taken into account. We hypothesize that here not only the presence of sun damage but also the absence of sun damage guided decisions. Hence, lesions that would have been diagnosed as pAK/BD can be diagnosed correctly as something else because of the lack of perilesional sun damage.

We expect our findings to be of specific importance for teledermatology. In contrast with live-consultations¹⁶ that are not always feasible, the assessment of chronic sun damage is not readily available in store-and-forward teledermatology applications, which frequently use close-up or dermatoscopy images without context information.

Limitations

Like in most publicly available dermatoscopic image data sets, individuals with darker skin type are underrepresented in our data set. This may lead to biased results.

The ratings from the reader study were collected in a gamified manner, which was similar but not identical to a real telemedical setting. In contrast to a real telemedical setting, no metadata were provided to participants. Metadata, such as age and anatomic site, could be surrogates for chronic sun damage and could improve the accuracy in a similar way as the information of perilesional sun damage. In practice, however, nearly any anatomic site can be prone to severe chronic sun damage. The association between anatomic site and pAK/BD is indirect, i.e. not causal, while there is a direct causal relation between UV-damage and pAK/BD. From two related predictors, one providing direct, the other indirect evidence, we selected the one with direct evidence.

References

- 1 Inskip M, Cameron A, Akay BN *et al*. Dermatoscopic features of pigmented intraepidermal carcinoma on the head and neck. *JDDG. Journal der Deutschen Dermatologischen Gesellschaft* 2020; **18**: 969–976.
- 2 Ackerman AB, Mones JM. Solar (actinic) keratosis is squamous cell carcinoma. *Br J Dermatol* 2006; **155**: 9–22.
- 3 Zalaudek I, Argenziano G, Leinweber B *et al*. Dermoscopy of Bowen's disease. *Br J Dermatol* 2004; **150**: 1112–1116.
- 4 Zalaudek I, Giacomel J, Schmid K *et al*. Dermoscopy of facial actinic keratosis, intraepidermal carcinoma, and invasive squamous cell carcinoma: a progression model. *J Am Acad Dermatol* 2012; **66**: 589–597.
- 5 Cameron A, Rosendahl C, Tschandl P, Riedl E, Kittler H. Dermoscopy of pigmented Bowen's disease. *J Am Acad Dermatol* 2010; **62**: 597–604.
- 6 Yang Y, Lin J, Fang S, Han S, Song Z. What's new in dermoscopy of Bowen's disease: two new dermoscopic signs and its differential diagnosis. *Int J Dermatol* 2017; **56**: 1022–1025.
- 7 Tschandl P, Rosendahl C, Kittler H. Dermoscopy of flat pigmented facial lesions. *J Eur Acad Dermatol Venereol* 2015; **29**: 120–127.
- 8 Lallas A, Lallas K, Tschandl P *et al*. The dermoscopic inverse approach significantly improves the accuracy of human readers for lentigo maligna diagnosis. *J Am Acad Dermatol* 2021; **84**: 381–389.

- 9 Tschandl P, Codella N, Akay BN *et al.* Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol* 2019; **20**: 938–947.
- 10 Bissoto A, Valle E, Avila S. Debiasing Skin Lesion Datasets and Models? Not So Fast. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 740–741.
- 11 Tschandl P, Rinner C, Apalla Z *et al.* Human–computer collaboration for skin cancer recognition. *Nat Med* 2020; **26**: 1229–1234.
- 12 Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 2018; **5**: 180161.
- 13 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>
- 14 Fink C, Blum A, Buhl T *et al.* Diagnostic performance of a deep learning convolutional neural network in the differentiation of combined naevi and melanomas. *J Eur Acad Dermatol Venereol* 2020; **34**: 1355–1361.
- 15 Muñoz-López C, Ramírez-Cornejo C, Marchetti MA *et al.* Performance of a deep neural network in teledermatology: a single-centre prospective diagnostic study. *J Eur Acad Dermatol Venereol* 2021; **35**: 546–553.
- 16 Andrees V, Klein TM, Augustin M, Otten M. Live interactive teledermatology compared to in-person care - a systematic review. *J Eur Acad Dermatol Venereol* 2020; **34**: 733–745.